The Block Relevance (BR) analysis supports the dominating effect of solutes hydrogen bond acidity on Δlog Poct–tol

European Journal of Pharmaceutical Sciences
Volume 53, 12 March 2014, Pages 50–54
DOI: 10.1016/j.ejps.2013.12.007

The Block Relevance (BR) analysis supports the dominating effect of solutes hydrogen bond acidity on $\Delta\log P_{oct\text{-}tol}$

Giuseppe Ermondi[1], Alessia Visconti[2], Roberto Esposito[2] and Giulia Caron[1]*

[1] *Molecular Biotechnology and Health Sciences Dept., Università degli Studi di Torino, via Quarello 15, 10135 Torino, Italy.*

[2] *Computer Science Dept., Università degli Studi di Torino, Corso Svizzera 185, 10149, Torino, Italy.*

E-mail: giulia.caron@unito.it

Telephone: +39 011 6708337

Keywords

BR analysis, Δlog P, lipophilicity, hydrogen bond acidity, VolSurf+.

1. Introduction

The role of hydrogen bond acidity, i.e., the ability of chemicals to act as hydrogen bond donors (HBD), is a crucial element in pharmaceutical sciences and medicinal chemistry. For instance, the counts of hydrogen bond donor and acceptor groups are part of the Ro5 parameters (Lipinski et al., 1997) used to predict drug-like properties and permeability; moreover, HBD count is a parameter in the CNS Multi-Parameter Optimization score (Wager et al., 2010).

Generally speaking, HBD groups are often important for increasing the binding to biological targets and water solubility. On the other hand they decrease membrane permeability and serve as a recognition feature for P-glycoprotein (P-gp), the efflux pump that excludes molecules from the brain.

The determination of hydrogen bond acidity is also important to determine the lipophilicity of drugs estimated from chromatographic measurements (Pallicer et al., 2012).

It has been shown that the difference between two log P values (Δlog P) obtained in different biphasic systems for example octanol/water and alkane/water ($\Delta\log P_{oct\text{-}alk} = \log P_{oct} - \log P_{alk}$), is informative of the solutes HBD properties (Abraham et al., 2010b) and thus useful in the prediction of drugs human fate (Liu et al., 2011). The recent interest for the application of $\Delta\log P_{oct\text{-}alk}$ in the drug discovery process stimulated the research for the design and implementation of tools for its prediction (Caron and Ermondi, 2005)(Toulmin et al., 2008)(Kenny et al., 2013).

Unfortunately the experimental determination of $\Delta\log P_{oct\text{-}alk}$ is strongly limited by the low alkane solubility of many compounds. To overcome this difficulty, toluene was proposed to replace alkanes (Zissimov et al., 2002)(Shalaeva et al., 2013) in lipophilicity measurements. The general idea is that log P_{tol} provides information similar to log P_{alk} but it is easier to assess experimentally. Indeed toluene has a better ability to solubilize organic compounds and still has a low alkane-like dieletric constant (ε), i.e., 2.38 (toluene) vs 2.02 (cyclohexane). According to these evidences $\Delta\log P_{oct\text{-}tol}$ is expected to be a convenient surrogate of $\Delta\log P_{oct\text{-}alk}$ for the determination of solutes HBD properties.

Here we use the Block Relevance (BR) analysis to describe the factorization of $\Delta\log P_{oct\text{-}tol}$ in its main components. BR analysis is a new tool that enables the mechanistic interpretation of PLS models (Ermondi

and Caron, 2012)(Caron and Ermondi, 2013). Also, we report about a BR analysis based comparison of $\Delta\log P_{oct\text{-}tol}$ and $\Delta\log P_{oct\text{-}alk}$.

In this study, we collected from the literature more than 200 experimental log P_{tol} values along with their corresponding log P_{oct} values. The dataset was processed using a purposely-built in-house software to remove molecules that are potentially able to form IMHBs. On the remaining structures the $\Delta\log P_{oct\text{-}tol}$ (= log P_{oct} - log P_{tol}) have been calculated and correlated with 82 VolSurf+ (VS+) descriptors through a PLS model. Finally the BR analysis has been used to group the 82 VS+ descriptors in six easy-to-interpret blocks and to graphically show the relevance of a certain block in the PLS model.

2. Methods

The log P_{tol} values used in this work have been retrieved from three different sources (Stephens et al., 2011) (Abraham et al., 2010a) (Box et al., 2012). Data obtained using DMSO as a cosolvent were discarded. In fact, it has been reported (Shalaeva et al., 2013) a small but systematic deviation in log P_{tol} values when DMSO is present. Log P_{oct} were also retrieved from literature, most from Abraham and coworkers (Abraham et al., 1994). $\Delta\log P_{oct\text{-}tol}$ were calculated as the difference log P_{oct} – log P_{tol}. The complete list of data can be found in Supporting Information (Table S1).

An in-house software was used to identify compounds with propensity to form intramolecular hydrogen bonds (IMHB) according to the topologies proposed by Kuhn and coworkers (Kuhn et al., 2010).

VS+ models were built by submitting the SMILES codes of the compounds to VS+ (version 1.0.7, http://www.moldiscovery.com) using default settings and four probes (OH2, DRY N1 and O probes that mimic respectively water, hydrophobic, HBA and HBD properties of the environment). PCA and PLS tools implemented in VS+ were used.

BR analysis was performed as described elsewhere (Ermondi and Caron, 2012)(Caron and Ermondi, 2013).

Processing was done on a two 8 cores Xeon E5 at 3.3GHz CPUs and 128GB of RAM.

3. Results and discussion

3.1. Dataset overview

We collected data for the log P_{tol}, log P_{oct} and Δlog $P_{oct-tol}$ values over 222 compounds. All the log Ps are referred to pH conditions of suppression of dissociation. These values span 6.5, 9.3 and 5.8 log P unities, respectively.

When VS+ processes the data, it associates each compound to the lipophilicity value of an "average" conformer built internally by an ad-hoc algorithm. In general terms this is a correct protocol because one can assume that the "average" conformer represents all conformers energetically accessible. However this assumption no longer holds when the molecule under study has strong propensity to form IMHBs since the molecule is then forced into a specific conformation. This latter could show a very different profile of VolSurf+ descriptors from the conformers without IMHBs.

To exemplify the influence of IMHB on Volsurf+ descriptors we selected ephedrine. Figure 1 shows two conformers of the drug. The conformer on the left forms IMHB, the reverse is true for the conformer on the right. The relative Boltzmann distribution of the two conformers can only be assessed by high level calculations which are beyond the scope of this study. Figure 1 shows the envelope which is accessible to, and interacts attractively with the DRY probe at -0.8 kcal/mol. The volume of the envelope corresponds to the descriptor D4. D4 values (9.25 and 2.125) vary considerably between the two conformers and thus the use of an "average" conformer is not advisable. Therefore, to produce a robust and unambiguous PLS model, we discarded 18 molecules showing topologies that are potentially able to form IMHBs. The compounds selected for removal are: 2,4-dinitrophenol, 2-nitroaniline, 2-nitrophenol, 8-hydroxyquinoline, atropine, desipramine, diclofenac, ephedrine, flumequine, fluoxetine, lidocaine, metoprolol, penbutolol, propranolol, quinine, salicylic acid, Sudan I (Z-form), and tramadol.

Insert Figure 1

The PCA analysis shows a good data dispersion (three PCs explain about 70% of the variance). More details are given in the Supporting Information (Annex S1).

3.2. PLS models

Experimental $\Delta\log P_{oct\text{-}tol}$ values were imported into VS+ as response variables (Y) and a relation between Y and the 82 VS+ descriptors (X) was sought using the PLS algorithm implemented in the software. The same procedure was applied to log P_{oct} and log P_{tol} for comparative purposes. The validation of the models was performed by means of an internal validation procedure (more details below in 3.2). Satisfactory statistical results have been obtained on all the models and reported in Table 1.

Insert Table 1

Correlations between calculated *vs* experimental values are shown in the Supporting Information (Fig. S1). As desired, slopes are close to 1 and intercepts close to 0. The lower quality of $\Delta\log P_{oct\text{-}tol}$ PLS model compared to log P models is probably related to the indirect determination of the descriptor.

The automatic identification of outliers is an open question in statistics and several excellent tests exist that deal with the problem (e.g., Grubbs' test and Rout method). In this work we chose to manually analyze deviant compounds. In our $\Delta\log P_{oct\text{-}tol}$ model a few compounds are potential outliers, most of them showing the same behavior also in log P models (e.g., biphenyl). Since these compunds have very high experimental log P values (> 3) some of the variance can be explained by detection limits affecting the accuracy of the experimental measurements obtained by shake-flask. Some others compounds show chemical substructures not well parametrised by VS+ descriptors (e.g., phenazopyridine). The exclusion of these compounds from the model does not produce significant improvement in the statistics (data not shown) and thus we prefer to retain them in the model.

We are aware (Ermondi and Caron, 2012) that some researchers in the QSAR field support internal validation, whereas others consider it as not sufficient for assessing the robustness of models and instead require an external validation (Wold and Sjostrom, 2001). In this study to obtain a correct estimation of the

real predictive ability of the model a Random Groups (RG) approach was used. The compounds in the data set were assigned in a random way to 4 groups, each one containing an equal (or nearly equal) number of objects. Then models were built keeping one of these groups out of the analysis until all of the objects were kept out once. The formation of the groups and the validation was repeated 20 times. Table 1 shows Q^2 for the RG procedure. Since all PLS models show R^2>0.6 and Q^2>0.5, they satisfy the Tropsha et al.'s validation rules. (Tropsha et al., 2003)

To further validate PLS models we finally tested the methodology on a dataset with a randomized Y order. As sought, the experiment produced unacceptable R^2 and Q^2 values (data not shown).

3.3. BR analysis

The mechanistic interpretation of a PLS model is generally obtained through the Variable Importance in the Projection (VIP) plot. A VIP plot displays VIP values as columns sorted in descending order with confidence intervals derived from jack-knifing. VIPs values are regarded as valuable tools in interpreting PLS models since they are able to take into account both the correlations with the target variable Y as well as the correlations within the X descriptors. However VIP plots (Supporting Information Fig. S2) are often hard to be interpreted. To overcome this problem we extended the VIPs analysis with the BR analysis, a recent methodology introduced in (Ermondi and Caron, 2012)(Caron and Ermondi, 2013).

BR analysis mandates the organization of the VS+ descriptors into six *blocks* (namely, Size, Water, DRY, N1, O and Others, definitions and more details in Table 2) which enables a straightforward understanding of the investigated phenomena (e.g., partitioning in the considered biphasic system). Indeed blocks provide an easy mechanistic interpretation based on the nature of the interaction of the solute with the environment represented by some tailored probes defined by the GRID methodology (Goodford, 1985)(Boobbyer et al., 1989)(Wade and Goodford, 1993). In practice one can apply to the BR analysis a mechanistic reasoning and easily compare different log P systems.

Insert Table 2

Graphical results of BR analysis are shown in Figure 2 which reports a pair of plots for each lipophilicity index (2A-B for log P_{oct}, 2C-D for log P_{tol} and 2E-F for Δlog $P_{oct\text{-}tol}$, respectively). Plots on the left indicate the relevance of each block in the model. Plots on the right split the contribution of each block into positive (BR (+)) and negative (BR (-)) components. BR (+) indicates how much the considered block favors solutes partitioning in the first phase whereas BR (-) indicates how much the considered block favors solutes partitioning in the second phase. A complete BR analysis includes inspection of both plots.

Insert Figure 2

For the sake of clarity we firstly discuss log P_{oct}. Figure 2A shows that *Size* is the most important block for this well-known molecular descriptor (about 32% of the weight of all blocks). Figure 2B indicates that the larger the molecule, the higher its partitioning in the octanol phase and thus the higher its log P_{oct}. The whole profile of intermolecular interactions is in line with BR plots obtained on different datasets (Caron and Ermondi, 2013).

BR plots for log P_{tol} are shown in Figure 2C and 2D. We notice a remarkable difference of these profiles with respect to log P_{oct} results revealing a more balanced contribution of *Size*, *Water*, *DRY* and *O* blocks (21%, 20%, 19% and 20%, respectively). The *O* block (solutes HBD properties, see Table 2) contribution varies considerably in log P_{oct} and log P_{tol} experiments: it is highly significant in Fig. 2C and favors the partitioning in the aqueous phase (as it is reported as mostly negative in Fig. 2D). These findings reflect the remarkable differences of toluene and octanol in their physico-chemical properties.

BR plots for Δlog $P_{oct\text{-}tol}$ are shown in Figure 2E-F and as expected are significantly different from those obtained for log Ps. The main block is the *O* block (Fig. 2E) that is positive in sign and represents about the 40% of the weight of all blocks. The two remaining polar blocks (*Water* and *N1*) are significant but less important (17% and 14%). Interestingly the *Size* and *DRY* blocks are not significant, confirming that the contribution of hydrophobicity to Δlog $P_{oct\text{-}tol}$ is negligible.

Very recently, Shalaeva and coworkers used $\Delta\log P_{oct\text{-}tol}$ to distinguish compounds able to form IMHB (samples) from those with similar structure but unable to do that (controls) (Shalaeva et al., 2013). Here we prove that $\Delta\log P_{oct\text{-}tol}$ mainly depends on the HBD properties of the solutes (see above) and thus support the application of $\Delta\log P_{oct\text{-}tol}$ in the intramolecular hydrogen bonding interpretation scheme. In fact a sample with strong propensity to form IMHB is expected to have lower exposure of HBD groups than its control. This results in a positive difference in $\Delta\log P_{oct\text{-}tol}$ between the control and the sample, as predicted for compound belonging to category I by Shalaeva et al. (Shalaeva et al., 2013).

3.4. $\Delta\log P_{oct\text{-}tol}$ *vs* $\Delta\log P_{oct\text{-}alk}$

As mentioned in the Introduction, $\Delta\log P_{oct\text{-}alk}$ is generally considered a valuable descriptor for HBD properties of solutes that has however a limited applicability due to the poor experimental accessibility of $\log P_{alk}$.

In a previous work, we used VS+ descriptors to build a PLS model on $\log P_{alk}$ values (Caron and Ermondi, 2005). In that study we built a dataset of 152 molecules and calculated the corresponding $\Delta\log P_{oct\text{-}alk}$ values. In the following we report about how these results compare with $\Delta\log P_{oct\text{-}tol}$ (204) results described in this paper.

For the same reasons that led us to discard compounds from the $\Delta\log P_{oct\text{-}tol}$ dataset (i.e., their propensity to form IMHBs), we had to discard 18 compounds from the $\Delta\log P_{oct\text{-}alk}$ dataset leaving us with 134 structures. Specifically we removed: 2-hydroxybenzoicacid, 2-nitroaniline, 2-nitrophenol, atenolol, carazolol, desipramine, diclofenac, ephedrine, flumequine, fluoxetine, lidocaine, metoprolol, penbutolol, phenazopyridine, propranolol, quinine, tramadol and warfarin.

The projection of the 134 compounds belonging to the $\Delta\log P_{oct\text{-}alk}$ dataset (empty circles) on the 2D PCA scores plot of the 204 molecule belonging to the $\Delta\log P_{oct\text{-}tol}$ dataset (full circles) is shown in Figure 3A. It shows that the two datasets cover similar chemical space regions. This result encouraged us to compare $\Delta\log P_{oct\text{-}alk}$ and $\Delta\log P_{oct\text{-}tol}$ BR analysis plots.

Insert Figure 3

Figure 3B emphasizes that $\Delta\log P_{oct\text{-}tol}$ depends to an higher degree on solutes HBD than $\Delta\log P_{oct\text{-}alk}$. In fact, the *O* block (solutes HBD properties) only represents about 20% of the weight of all blocks in the model based on $\Delta\log P_{oct\text{-}alk}$ values, while it represent about 40% of the weights in the $\Delta\log P_{oct\text{-}tol}$ based model (see Section 3.3).

4. Conclusions

In this paper we showed that $\Delta\log P_{oct\text{-}tol}$ strongly depends on HBD of solutes. This was proven with a recently developed computational tool based on VS+ descriptors and PLS algorithm, i.e., the Block Relevance analysis. Our findings support the use of $\Delta\log P_{oct\text{-}tol}$ as a molecular descriptor for the determination of the solutes HBD properties in drug discovery. In particular this study supports the recently published IMHB interpretation scheme (Shalaeva et al., 2013) that uses $\Delta\log P_{oct\text{-}tol}$ as a tool to distinguish compounds based on their propensity to form IMHBs.

Acknowledgements

This work has been supported by Ateneo Compagnia di San Paolo-2012-Call 2, LIMPET project.

References

Abraham, M.H., Chadha, H.S., Whiting, G.S., Mitchell, R.C., 1994. Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the Dlog P parameter of Seiler, J. Pharm Sci. 83, 1085-100.

Abraham, M.H., Acree, W.E., Leo, A.J., Hoekman, D., Cavanaugh, J.E., 2010a. Water – Solvent Partition Coefficients and D Log P Values as Predictors for Blood – Brain Distribution; Application of the Akaike Information Criterion, J. Pharm. Sci., 99, 2492–2501.

Abraham, M.H., Smith, R.E., Luchtefeld, R.O.N., Boorem, A.J., Luo, R., Acree, W.E, 2010b. Prediction of Solubility of Drugs and Other Compounds in Organic Solvents. J Pharm Sci., 99, 1500-15.

Boobbyer, D.N.A., Goodford, P.J., Mcwhinnie, P.M., Wade, R.C., 1989. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure, J. Med. Chem. 32, 1083–1094.

Box, K.J., Comer, J., Ruiz, R., Mole, J., Frake, L., 2012. Determination of hydrogen bonding properties using log P measurements in different solvent-water systems, in: 2012 AAPS Annual Meeting and Exposition. Poster N°M1044.

Caron, G., Ermondi, G., 2005. Calculating virtual log P in the alkane/water system (log P(N)(alk)) and its derived parameters delta log P(N)(oct-alk) and log D(pH)(alk). J. Med. Chem. 48, 3269–79.

Caron, G., Vallaro, M., Ermondi, G., 2013. The Block Relevance (BR) analysis to aid medicinal chemists to determine and interpret lipophilicity. Med.Chem.Comm. doi:10.1039/c3md00140g.

Ermondi, G., Caron, G., 2012. Molecular interaction fields based descriptors to interpret and compare chromatographic indexes. J. Chromatogr. A 1252, 84–9.

Goodford, P.J., A, 1985. computational procedure for determining energetically favorable binding sites on biologically important macromolecules, J. Med. Chem. 28, 849–857.

Kenny, P.W., Montanari, C., Prokopczyk, I.M., 2013. ClogPalk: a method for predicting alkane/water partition coefficient. Journal of computer-aided molecular design J. Comput. Aided Mol. Des. 27, 389–402.

Kuhn, B., Mohr, P., Stahl, M., 2010. Intramolecular hydrogen bonding in medicinal chemistry. J. Med. Chem. 53, 2601–11.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 23, 3–25.

Liu, X., Testa, B., Fahr, A., 2011. Lipophilicity and its relationship with passive drug permeation. Pharm.Res. 28, 962–77.

Pallicer, J.M., Pascual, R., Port, A., Rosés, M., Ràfols, C., Bosch, E., 2012. The contribution of the hydrogen bond acidity on the lipophilicity of drugs estimated from chromatographic measurements. Eur. J. Pharm. Sci. 48, 484–493.

Shalaeva, M., Caron, G., Abramov, Y.A., O'Connell, T.N., Plummer, M.S., Yalamanchi, G., Farley, K.A., Goetz, G.H., Philippe, L., Shapiro, M.J., 2013. Integrating intramolecular hydrogen bonding (IMHB) considerations in drug discovery using ΔlogP as a tool. J.Med.Chem. 56, 4870−4879.

Stephens, T.W., Loera, M., Quay, A.N., Chou, V., Shen, C., Wilson, A., Acree, W.E., Abraham, M.H., 2011. Correlation of Solute Transfer Into Toluene and Ethylbenzene from Water and from the Gas Phase Based on the Abraham Model. The Open Thermodynamics Journal, 5, 104-121.

Toulmin, A., Wood, J.M., Kenny, P.W., 2008. Toward prediction of alkane/water partition coefficients. J.Med.Chem. 51, 3720–30.

Tropsha, A., Gramatica, P., Gombar, V.K., 2003. The Importance of Being Earnest : Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. QSAR Comb. Sci. 22, 69–77.

Wade, R.C., Goodford, P.J., 1993. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. J. Med. Chem 36, 148–156.

Wager, T.T., Hou, X., Verhoest, P.R., Villalobos, A., 2010. Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. ACS Chem Neurosci 1, 435–49.

Wold, S., Sjostrom, M., Eriksson, L. 2001. PLS-regression : a basic tool of chemometrics . Chemometr. Intell. Lab. 58, 109–130

Zissimos, A.M., Abraham, M.H., Barker, M.C., Box, K.J., Tam, K.Y., 2002. Calculation of Abraham descriptors from solvent–water partition coefficients in four different systems; evaluation of different methods of calculation. J. Chem. Soc., Perkin Trans. 2, 470-477.

Figure 1. Ephedrine: MIFs generated by the DRY probe at −0.8 kcal/mol in the presence (left) and in the absence (right) of IMHBs. The volume of the envelope corresponds to the descriptor D4. D4 values are 9.25 and 2.125, respectively.

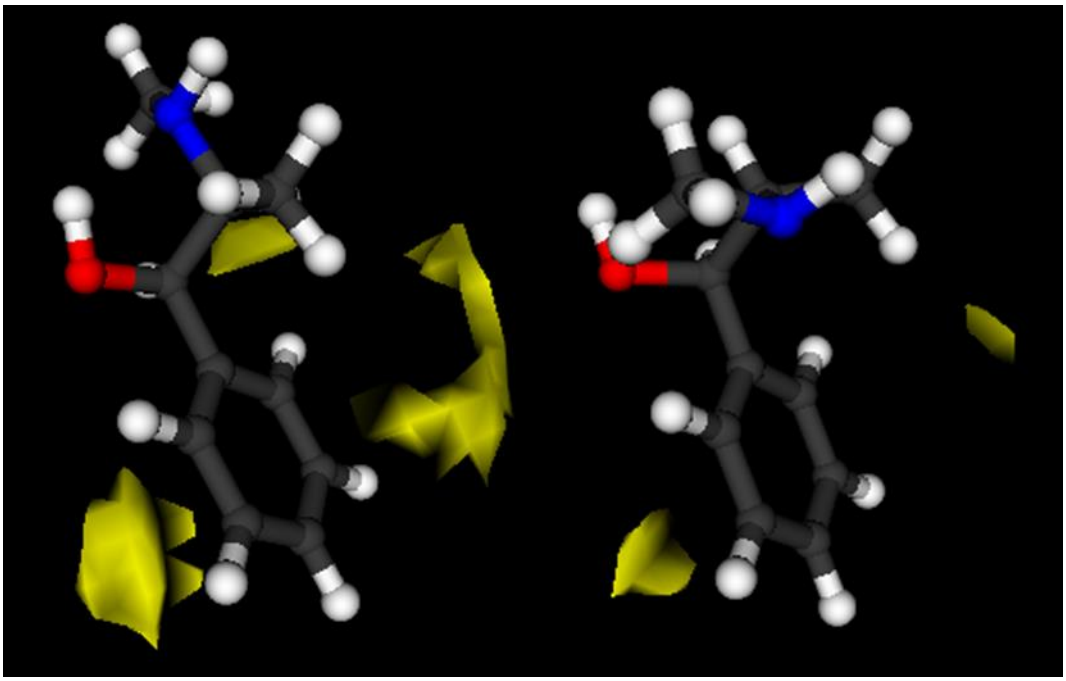

Figure 2. BR analysis graphical outputs. A color code is associated to each block: green for Size, cyan for OH2 (Water), yellow for DRY, blue for N1, red for O and grey for Others. Subfigures A and B reports results about log P_{oct}; subfigures C and D reports about log P_{tol}; subfigures E and F reports about Δlog $P_{oct\text{-}tol}$. Plot on the left (A, C and E) indicate the relevance of the block in the model. Plot on the right (B, D and F) splits the contribution of each block in positive (BR (+)) and negative (BR (-)) components. The dotted line shows the blocks significance threshold.

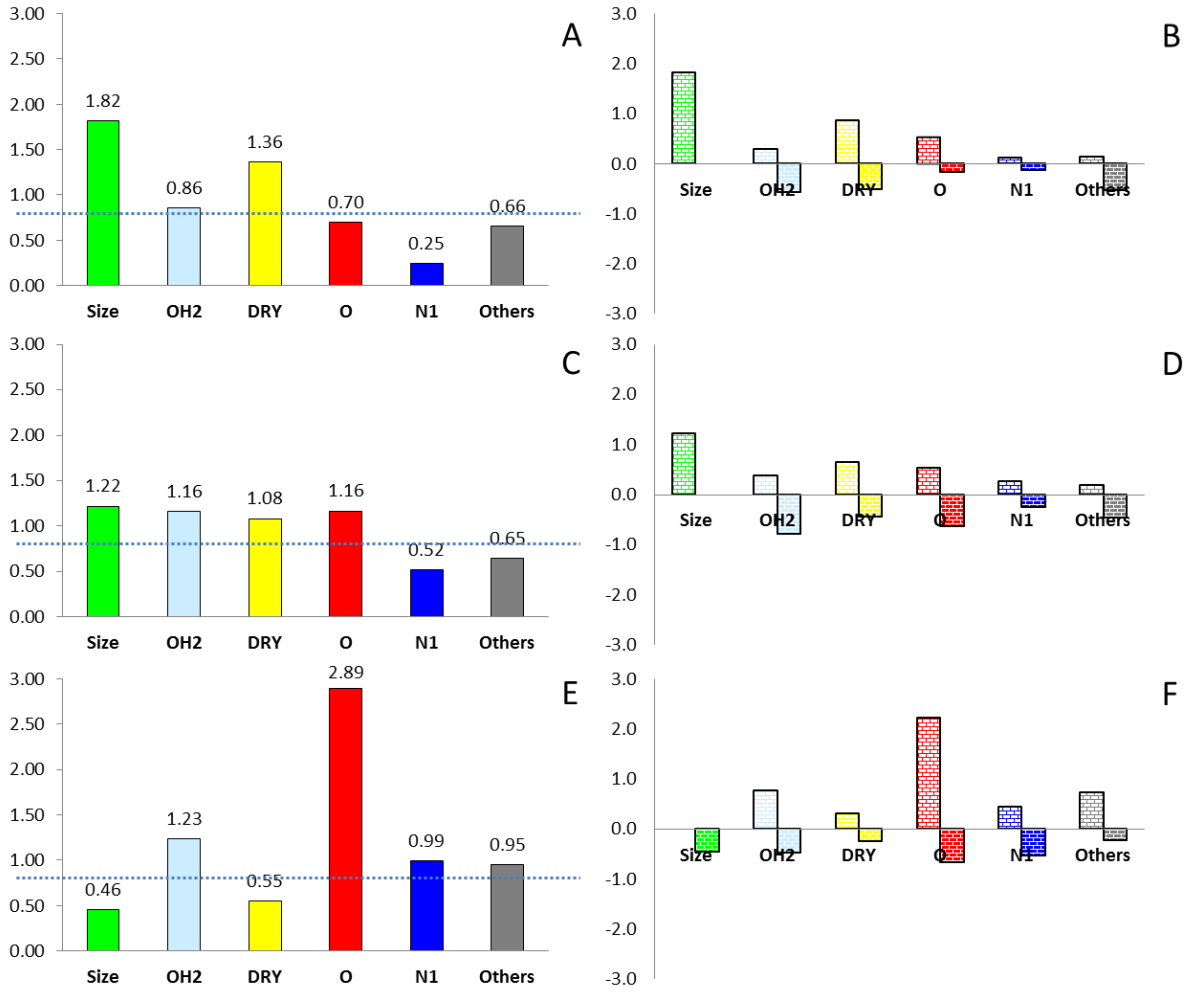

Figure 3. Δlog $P_{oct\text{-}alk}$ data analysis: A) The projection of the 134 compounds belonging to the Δlog $P_{oct\text{-}alk}$ dataset (empty circles) on the 2D PCA scores plot of the 204 molecule belonging to the Δlog $P_{oct\text{-}tol}$ dataset (full circles); B) BR analysis graphical output for Δlog $P_{oct\text{-}alk}$

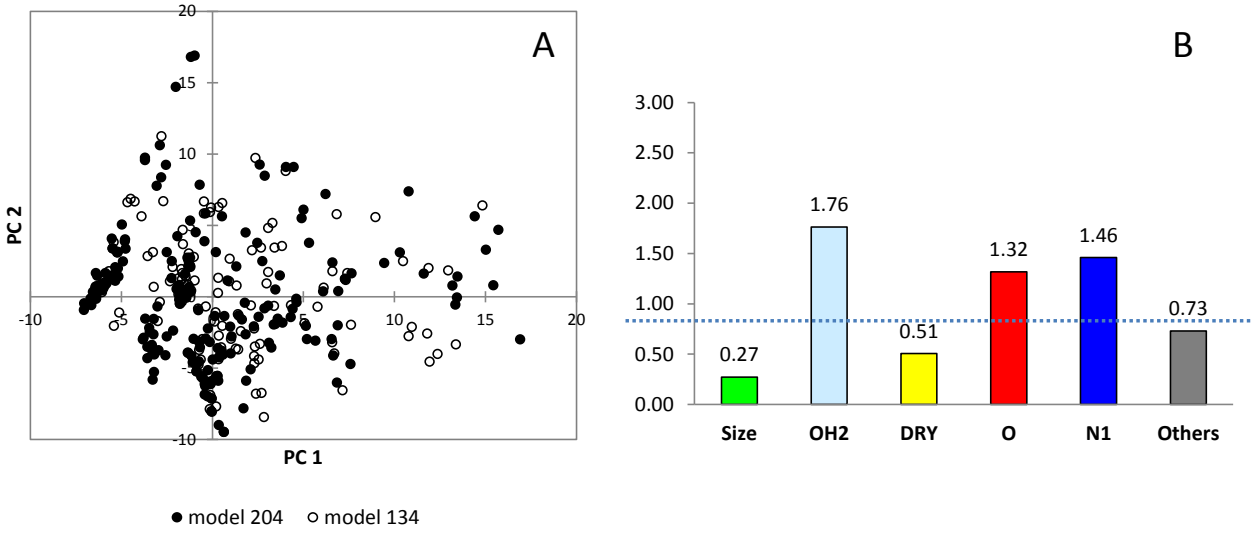

Table 1. PLS models (N = number of observations, R^2 = cumulative determination coefficient, Q^2(RG) = cross-validated correlation coefficient (see text for details), LV = number of latent variables, RMSE = root mean square of the errors).

System		N	R^2	Q^2(RG)	LV	RMSE
$\Delta\log P_{oct\text{-}tol}$	**82**	**204**	**0.74**	**0.56**	**5**	**0.62**
$\log P_{oct}$	82	204	0.80	0.68	5	0.76
$\log P_{tol}$	82	204	0.85	0.75	5	0.92

Table 2. Block definition. A color code is associated to each block: green for Size, cyan for OH2 (Water), yellow for DRY, blue for N1, red for O, and grey for Others.

Block	Definition	Color code
Size	descriptors that characterize the size and shape of the solute	green
OH2 (Water)	descriptors that express the solute's interaction with water molecules (= with the GRID OH2 probe)	light blue
*N1**	descriptors that describe the solute's ability to form hydrogen bond interactions with the GRID N1 probe (that mimics the system); roughly superposable with Abraham's $\Sigma\beta^{H}_{2}$	blue
*O**	descriptors expressing the solute's ability to form hydrogen bond interactions with the GRID O probe (that mimics the system); roughly superposable with Abraham's $\Sigma\alpha^{H}_{2}$	red
DRY	descriptors describing the solute's propensity of the solute to participate in hydrophobic (= with the GRID probe DRY) interactions	yellow
Others	descriptors mainly describing the imbalance between hydrophilic and hydrophobic regions	grey

** For the sake of clarity, to identify hydrogen bonding (HB) interactions, i.e., Hydrogen Bond Acceptor capability (HBA) and Hydrogen Bond Donor capability (HBD), we refer to the probe's properties and not to the solute (see following scheme).*